

Satellite monitoring system to determine critical regions in a terrestrial transport route by applying supervised learning techniques

Karim Guevara Puente de la Vega, Phd.^{1,2}, Miguel Angel Mamani Zeballos²

¹Universidad Nacional de San Agustín, Arequipa-Perú, kguevarap@unsa.edu.pe

²Universidad Católica de Santa María, Arequipa-Perú, kguevara@ucsm.edu.pe, 43108106@ucsm.edu.pe

Abstract– At present, the use of satellite monitoring systems involves much more than the storage of historical data. Because with proper processing, these large volumes of data can provide new, relevant, useful information of great importance to transport companies. Allowing companies to have information on the incidents in the vehicular displacement, to assess the level of security of the transport services provided. For this, it is necessary that all existing incidents or risks are counted and / or detected. In the present article a satellite monitoring system is presented for the determination of critical regions in terrestrial transport routes, the same one that was conceived from the use of supervised learning techniques. The analysis carried out starts from the structure of the standardized historical reports that correspond to the monitoring of vehicular units. These reports were generated by a current system of consultation and monitoring in real time. Therefore, through the use of these techniques it was possible to achieve the discovery of knowledge stored in databases.

Key words - Algorithm NN, Discovery of knowledge in databases, Data mining, Vehicle monitoring, Rules of Decision.

I. INTRODUCCIÓN

En el Perú, el servicio de transporte terrestre interprovincial es normado, supervisado, fiscalizado y sancionado por la Superintendencia de Transporte Terrestre de Personas, Carga y Mercancías (SUTRAN); entidad creada mediante la Ley N° 29380 [15], adscrita al Ministerio de Transportes y Comunicaciones (MTC). La SUTRAN cuenta con un Centro de Gestión y Monitoreo [8], para incrementar el nivel de cumplimiento de las normas por parte de los operadores del servicio de transporte terrestre, dichas funciones son respaldadas por el Decreto Supremo N° 017-2009-MTC, en el cual se estipula que las unidades vehiculares de las empresas destinadas al transporte terrestre, deben contar con un sistema de control y monitoreo inalámbrico que transmita mediante un dispositivo GPS a la autoridad en forma permanente la información del vehículo en ruta.

Por lo tanto, todas las transmisiones están en un almacén de datos y pueden ser visualizadas mediante un reporte con filtros de intervalo de fechas y las placas vehiculares registradas; a pesar de ello, todos los servicios de transporte operados por una determinada empresa no pueden ser visualizados de manera distintiva, debido al flujo de trabajo limitado en la mayoría de sistemas empleados en los

terminales terrestres, y siendo, una de las necesidades primordiales de las empresas de transporte realizar el registro de los viajes de sus unidades vehiculares. El nivel de interacción con los sistemas de monitoreo disponibles es muy bajo, y una de sus funcionalidades principales actualmente es la de permitir visualizar datos históricos de los distintos estados de la geolocalización en el intervalo de tiempo de los vehículos consultados.

Hoy en día, el tema de calidad de servicio no es ajeno al rubro de transportes, para ello es necesario evaluar el desplazamiento de un vehículo a nivel de viajes; sin embargo en la actualidad, las distintas transmisiones solo son evaluadas para verificar si un vehículo cometió una infracción por exceso de velocidad y un viaje es tema de investigación cuando se pretende esclarecer las causas de un accidente de tránsito.

En consecuencia, resulta muy necesario tener en cuenta que los sistemas de monitoreo satelital vigentes en los terminales terrestres deben ser mejorados o reemplazados, lo que involucra procesos de identificación de los problemas, oportunidades y el cumplimiento de una serie de objetivos propuestos. Es así que la influencia de incorporar las tecnologías de información puede crecer exponencialmente ante la exigencia de competitividad en el mercado; para resolver problemas específicos previamente identificados.

Dentro de los riesgos existentes en la actividad de transporte terrestre de personas, mercancías o materiales peligrosos; los conductores constituyen o representan uno de los riesgos más críticos a controlar o reducir. Para ello, el sistema propuesto describe un flujo de trabajo más completo que permite evaluar y controlar el desempeño de los conductores, a través de la determinación de las regiones críticas en las rutas de transporte terrestre en el Perú.

El resto del artículo está organizado de la siguiente manera: la sección 2 describe el cuerpo del conocimiento que apoyó en el desarrollo e implementación del sistema. En la sección 3 se precisa el problema que se intenta resolver con la implementación del sistema de monitoreo satelital. En la sección 4 se presenta aspectos de la implementación del sistema propuesto basado en las fases de la metodología CRISP-DM. En la sección 5 se describe los resultados obtenidos por el sistema propuesto respecto a los obtenidos por el sistema vigente. Finalmente se detalla conclusiones

Digital Object Identifier (DOI): <http://dx.doi.org/10.18687/LACCEI2017.1.1.205>
ISBN: 978-0-9993443-0-9

ISSN: 2414-6390

15th LACCEI International Multi-Conference for Engineering, Education, and Technology: “Global Partnerships for Development and Engineering Education”, 19-21 July 2017, Boca Raton FL, United States.

II. CUERPO DEL CONOCIMIENTO

A. Trabajos previos

En esta sección, se describen una serie de propuestas y medidas de control formuladas en documentos y publicaciones, que evidencian la preocupación ante la problemática persistente referente a la seguridad vial y reducción de accidentes en el tránsito vehicular.

En el trabajo presentado por García et al. [1], se exponen las diversas posibilidades de aplicación a casos prácticos referentes a los GPS, hoy en día muchas empresas ya disponen de información almacenada de varios meses e incluso años; asimismo, presentan un enfoque para determinar las velocidades para cada uno de los segmentos de vía que componen la red vial de una ciudad, y a partir de las velocidades registradas se realiza un análisis para el desplazamiento promedio estimado de vehículos en un determinado día.

Sarmiento [2] en su trabajo desarrolla un prototipo de sistema de información geográfico en entorno web que permite detectar los eventos ocurridos en ruta de vehículos con GPS de una forma interactiva y funcional para el usuario, mediante la integración de las tecnologías de bases de datos, cache de mapas y y AJAX; el diseño del sistema corresponde al patrón de diseño MVC para permitir manipular y visualizar los datos geográficos procesados con bases de datos espaciales.

En el trabajo de Basulto [3] se describe como obtener o descubrir conocimiento acerca de los accidentes de tránsito que afectan a una empresa de transporte en un determinado periodo de años mediante la aplicación de técnicas de minería de datos, este trabajo fue posible mediante el acceso a una base de datos de las afectaciones ocurridas en dicha empresa con datos dispersos y sin normalizar; por medio de una metodología se analizan los accidentes de tránsito a fin de facilitar la extracción de la información implícita en los datos de una manera organizada de lo general a lo particular, para ayudar a que se adopten medidas encaminadas a reducir el nivel de accidentalidad.

B. GPS (Global Position System)

GPS es un sistema que tiene como objetivo la determinación de las coordenadas espaciales de puntos respecto de un sistema de referencia mundial. Los puntos pueden estar ubicados en cualquier lugar del planeta, pueden permanecer estáticos o en movimiento y las observaciones pueden realizarse en cualquier momento del día [4].

C. Metodología CRISP-DM

CRISP-DM (Cross - Industry Standard Process for Data Mining) describe los enfoques de uso común que los expertos

usan para enfrentar a los problemas de minería de datos [5]. Este modelo de proceso proporciona una visión general del ciclo de vida de un proyecto de minería de datos dividido en seis fases, como se aprecia en la fig. 1.

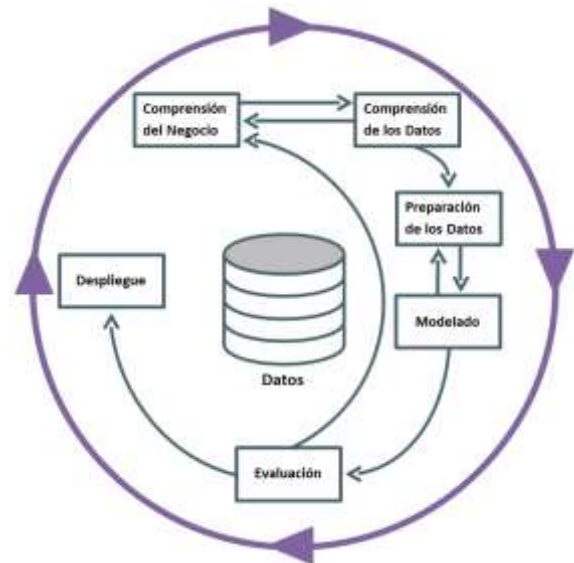


Fig. 1 Fases de la metodología CRISP-DM. [15]

Esta metodología percibe diferentes objetivos entre los principales se puede encontrar:

- Perseguir el cumplimiento de objetivos desde un punto de vista empresarial, por ende da preferencia a la comprensión del negocio.
- Desarrollar proyectos de minería de datos mediante un proceso estandarizado.
- Minimizar los costos que implica un proyecto de minería de datos en las empresas.

III. DESCRIPCIÓN DEL PROBLEMA

En el Perú, el servicio de transporte terrestre interprovincial para personas, es controlado y monitoreado mediante los sistemas de localización automática de vehículos en línea; este procedimiento establecido desde hace ya varios años atrás mediante el Decreto Supremo N° 017-2009-MTC [6], tiene como finalidad brindar mayor seguridad a los usuarios del servicio de transporte. A pesar de esta medida de control adicional, el porcentaje de accidentes de tránsito registrados anualmente [7], no ha mostrado una disminución considerable, inclusive si se toman en cuenta como referencia las estadísticas de años anteriores a la implementación de dicha medida.

El desarrollo adecuado de las funciones del Centro de Gestión y Monitoreo [8], está respaldado por los sistemas de las empresas de monitoreo vehicular autorizadas por el MTC. Las distintas transmisiones del desplazamiento vehicular

monitoreado, pueden ser exportadas mediante un tipo de Reporte de Tracking de Unidades [9], mediante estos reportes se puede apreciar que dichos sistemas poseen una funcionalidad limitada debido a que no es factible tener la certeza del número total de servicios de transporte realizados por un vehículo en un rango de fechas determinado (e.g., días, semanas, meses); en estas condiciones, los distintos tramos monitoreados sólo representan un historial de transmisiones, y en la actualidad resulta una necesidad fundamental contar con información íntegra y confiable a la hora de realizar o evaluar el nivel de seguridad del servicio de transporte prestado por una empresa, independientemente de los tipos de calidad de servicios brindados por las distintas empresas.

IV. METODOLOGÍA

El sistema propuesto y descrito en este artículo, está basado en torno a las pautas detalladas en el trabajo de García et al. [1], debido a que las reglas de decisión están consideradas en base a los límites de velocidad permisibles para una determinada zona de desplazamiento vehicular. En esta propuesta se realiza un control de los distintos eventos ocurridos en geozonas registradas y delimitadas en forma de polígonos para evaluar el ingreso o salida a las zonas permitidas (Geocercas) pertenecientes a la ruta asignada a un vehículo. Para la propuesta se considera la idea de cómo funcionan las geocercas pero no se las utiliza, sino más bien se hace uso de un gestor de base de datos normal que permita contener el flujo propuesto y así almacenar los datos de salida esperados; dicha información puede servir como apoyo en la toma de decisiones a nivel estratégico y será exclusivamente de uso corporativo; para ello, el sistema propuesto contiene la funcionalidad de administrar en forma más ordenada las transmisiones del desplazamiento vehicular, organizándolos en viajes realizados. El trabajo de Basulto [3], es el antecedente que expone la viabilidad de analizar desde otro punto de vista las posibles causas de accidentes en los datos históricos; por lo que en la propuesta se realiza un análisis de lo particular a lo general; debido a que un registro de los datos de monitoreo de un vehículo en movimiento corresponde a un instante de tiempo único con atributos de velocidad y geolocalización muy exactos, para referenciar con más exactitud el tipo de geozona a la cual pertenece la señal de GPS registrada forma parte de uno de los desafíos a superar en el desarrollo del sistema de monitoreo satelital presentado; por lo que es de vital importancia contar con un conjunto de entrenamiento de geopuntos lo bien definidos que abarque un gran porcentaje de la ruta a evaluar.

Tomando en cuenta los detalles mencionados anteriormente y considerando las especificaciones y recomendaciones establecidas en el “Texto Único Ordenado del Reglamento Nacional de Tránsito”, detallado en el Decreto Supremo N° 016-2009-MTC [10], para los límites de velocidad permisibles en cada uno de los distintos geopuntos

(conjunto de entrenamiento) asignados a una ruta, fue necesario organizar, planificar, evaluar y desarrollar los algoritmos que permitan la realización de la clasificación de correspondencia de los datos históricos; permitiendo de esta manera aplicar la minería de datos además de alcanzar los objetivos propuestos.

Asimismo, este sistema de monitoreo satelital nace como parte de una de las mejoras necesarias para los sistemas de monitoreo vehicular satelital empleados por las empresas de transporte en los terminales terrestres en el Perú. Para esto se utilizó la metodología CRISP-DM, descrita en la sección anterior, teniendo en cuenta las distintas fases que propone para realizar el proceso de minería de datos aplicado a este problema, lo que se describe a continuación.

A. Fase I: Comprensión del negocio.

El objetivo principal del negocio de este tipo de empresas es brindar un servicio de calidad que se destaque por puntualidad y seguridad, respaldado por la no existencia de accidentes de tránsito en los cuales recaiga la responsabilidad en la empresa de transporte. Para dar cumplimiento a este objetivo los operadores tienen que respetar y seguir la filosofía de “cero incidentes en ruta”. Por lo tanto, se debe reducir o eliminar el número de incidentes registrados en un viaje, y para lograrlo, tienen que estar identificados todos o la mayoría de dichos incidentes. Teniendo en cuenta que el límite máximo permisible de velocidad no es realmente 90 km/h, debido al sustento contenido en el Informe N° 224-2013/GEL-INDECOPI [11], por el cual a fin de no generar sanciones incorrectamente impuestas, se procede a considerar un margen de error de 10 km/h en la velocidad detectada del vehículo que se encuentra en movimiento lo que se norma a través del Decreto Supremo N° 009-2015-MTC [12]. En el Perú, las empresas encargadas de brindar el servicio de transporte terrestre, están obligadas a adquirir el servicio de monitoreo satelital vehicular para todas sus unidades de transporte según el Decreto Supremo N° 017-2009-MTC [6], el cual transmita en forma permanente la información del vehículo en ruta al Centro de Gestión y Monitoreo [8] de la SUTRAN. Por su parte la SUTRAN es responsable de identificar y sancionar los excesos de velocidad de los vehículos infractores de las distintas empresas de transporte terrestre. Por otra parte, los sistemas de monitoreo satelital vigentes no cuentan con los recursos necesarios para gestionar una interpretación ideal de los distintos eventos registrados en el desarrollo de uno o varios viajes monitoreados por el dispositivo GPS asignado a una unidad de transporte. La función principal de dicho dispositivo simplemente está limitada a transmitir las coordenadas de geolocalización (latitud y longitud) con un margen de error de 5 a 15 metros, velocidad instantánea de desplazamiento (Km/h), la fecha y la hora forman parte del vector de datos, que genera un registro histórico de la ruta cubierta por el vehículo monitoreado.

El propósito del sistema de monitoreo satelital que se presenta en este artículo se enfoca en descubrir los distintos incidentes ocultos registrados en un determinado viaje realizado por una unidad de transporte; los factores, exceso de velocidad y desacato de señales son determinantes en la detección de incidentes tal y como se muestra en la fig. 2 la cual representa el reporte estadístico realizado y actualizado hasta el año 2015 en base a los accidentes de tránsito fatales y no fatales declarados a nivel nacional en las unidades de la PNP, Accidentes Declarados en las Unidades de la PNP [7].

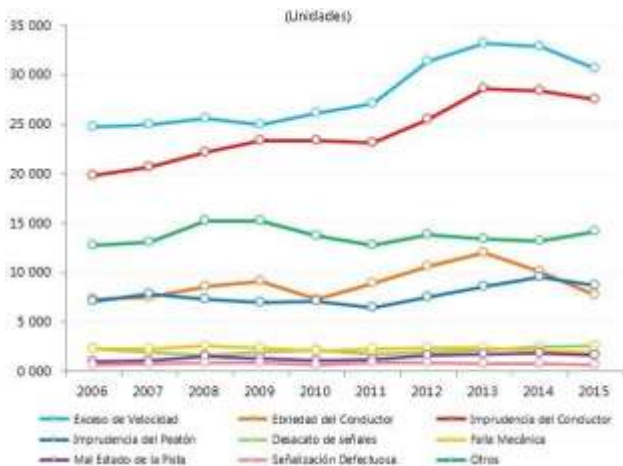


Fig. 2. Evolución de los Accidentes de Tránsito Fatales y No Fatales, según causa: 2006-2015. [7]

Actualmente, la mayoría de sistemas utilizados por las empresas de transporte terrestre a nivel nacional, más allá de visualizar la última ubicación de todas sus unidades monitoreadas, por lo general, emplean dichos sistemas para realizar la consulta del estado actual de una determinada unidad de transporte; y solo en calidad de investigación o control, realizan un post-monitoreo, para confrontar dicho desplazamiento con la hoja de ruta predeterminedada y personalizada según la política de la empresa. Sin embargo, al pretender realizar la distinción del recorrido vehicular registrado entre una serie de n viajes entre dos ciudades o más, puede terminar siendo una situación cuestionable, incómoda y/o engorrosa para los usuarios finales.

La solución planteada a este problema está basada en el uso del proceso de descubrimiento de conocimiento en bases de datos o KDD, y consiste, en aplicar el algoritmo de clasificación por vecindad en conjunto con las reglas de decisión obtenidas por el algoritmo de aprendizaje de árboles. La obtención de resultados ideales o aceptables dependerá del nivel de preparación e integridad de los datos a procesar con los algoritmos descritos más adelante, los mismo que en conjunto con las muestras de entrenamiento a emplear, permitirán obtener el vector final a evaluar mediante las reglas de decisión que pueden permitir la obtención de información oculta, predecible e importante.

B. Fase 2: Comprensión de los datos.

La recolección de los datos iniciales a clasificar, es posible mediante el módulo de reportes del sistema de monitoreo vigente de la empresa contratada para brindar dicho servicio; este módulo, denominado “Reporte de Tracking de Unidades”, según un determinado rango de fechas asignado para una o varias unidades de transporte establecido por el usuario, genera un archivo exportado en el formato de CSV. Este reporte, contiene las distintas transmisiones existentes pertenecientes a los intervalos de tiempo en que fueron activados los dispositivos de monitoreo y no corresponden necesariamente a la cobertura de una ruta y poseen campos tales como: fecha, hora, latitud, longitud, velocidad (Km/h), odómetro (km) y referencia del lugar en que fue procesada cada una de las transmisiones registradas en intervalos discontinuos de tiempo según la programación, el modelo o el nivel de señal del dispositivo GPS.

La estructura de los datos iniciales está contenida y distribuida en el *Reporte de Tracking de Unidades (2015)* [9], y los campos que incluyen son: nombre del cliente, placa de la unidad de transporte, fecha y hora de la transmisión, fecha y hora de envío de la transmisión, latitud de la transmisión, longitud de la transmisión, velocidad instantánea enviada en la transmisión, odómetro, lugar de referencia de la transmisión, y el total de transmisiones existentes en el archivo origen. Cada uno de los datos registrados en las “ n transmisiones” del dispositivo GPS conforma un vector de datos 100% completo debido a que los distintos dispositivos están homologados y estandarizados; además de contar con una precisión muy exacta en cuanto a la ubicación de geoposicionamiento, por lo tanto, la presencia de datos outlier es nula. La calidad de los datos es muy buena debido a que no existen valores perdidos y el formato de fechas está unificado para todos los registros.

Al explorar más detenidamente los datos iniciales, se aprecia que existen campos que son irrelevantes (e.g., el campo que indica el número total de transmisiones existentes en el archivo origen en forma repetitiva para todos los registros); otros campos necesitan una conversión de nominal a numérico, debido a que en su estado original no se pueden procesar numéricamente (las coordenadas de posicionamiento deben estar en números decimales); por lo señalado, fue necesario realizar un proceso de transformación de datos, el mismo que permitió la generación de nuevos campos que no eran considerados en el sistema de monitoreo satelital vigente; considerando que ya son más de 6 años a la actualidad que se espera más de una mejora en dicho servicio ofertado en el Perú.

C. Fase 3: Preparación de los datos.

En vista que la preparación de los datos es fundamental para realizar una óptima minería de datos, se definió el conjunto de vecinos más cercanos, el conjunto de

entrenamiento de reglas de decisión, los reportes recopilados correspondientes a los viajes a evaluar, para habilitar el almacén de datos, se realizó el proceso de ETL para importar los registros de monitoreo necesarios y se integró con la ayuda del algoritmo NN el vector de datos necesario para aplicar las reglas de decisión.

Debido a que no fue factible acceder a la base de datos del sistema de monitoreo vigente, por temas relacionados con la seguridad; resultó necesario crear el almacén de datos a partir de los reportes de tracking de unidades exportados por el sistema de monitoreo satelital vigente para simular la recepción de transmisiones GPS. Para que la importación de datos sea exitosa, el sistema propuesto cuenta con un módulo encargado del proceso ETL. Este proceso realiza la migración de los datos a clasificar previamente transformados con un algoritmo de integración de datos, el mismo que se encargará principalmente de eliminar campos irrelevantes y crear o modificar los valores o campos para una adaptación ideal de los algoritmos complementarios.

El diseño necesario de la DB que soporta el flujo de trabajo para el sistema propuesto, cuenta con las siguientes tablas: transmisiones, vehículos, geopuntos, rutas, viajes, trabajadores, cargos y usuarios.

D. Fase 4: Modelado.

Para llevar a cabo esta fase se empleó las técnicas de minería de datos predictivas a través de las cuales se especificó el modelo para los datos en base a un conocimiento teórico previo. El modelo supuesto para los datos se debe contrastar después del proceso de minería de datos para poder ser aceptado como válido.

El modelo de clasificación empleado en el sistema propuesto utiliza dos algoritmos: el algoritmo NN que realiza un análisis discriminante que permite asignar o clasificar nuevos individuos dentro de grupos previamente definidos, y el algoritmo ID3 que permite realizar la codificación de las reglas de decisiones a través de las cuales es posible clasificar los datos en grupos basados en los valores de las variables.

Para que el *algoritmo NN* realice la clasificación más acertada de correspondencia, depende directamente del tamaño del universo del conjunto de entrenamiento o geopuntos registrados y asociados a una o varias rutas; sería ideal utilizar un conjunto de entrenamiento de similar tamaño para generar las reglas de decisión a partir de datos relacionados a los accidentes de tránsito registrados en el Perú, pero actualmente esa es una de las debilidades a afrontar según el Plan Nacional de Seguridad Vial 2015-2024 desarrollado por el Consejo Nacional de Seguridad Vial [13], el mismo que detalla que existe una deficiente recopilación de datos de accidentes de tránsito, así como un déficit en la calidad de datos recopilados de accidentes de tránsito. Ante

esta situación; el sistema propuesto, utiliza un conjunto de entrenamiento relativamente pequeño para obtener las reglas de decisión; a pesar de ello, ese conjunto considera todas las combinaciones lógicas posibles en las cuales pueda existir riesgo.

Finalmente, para diseñar el modelo de clasificación más adecuado se tomó en cuenta objetivo principal de este trabajo que es la determinación de las regiones críticas en una ruta de transporte terrestre, y para esto, como ya se mencionó antes, se utilizó un conjunto de técnicas de aprendizaje supervisado cuyas características describimos a continuación así como su contribución al diseño del sistema propuesto.

- Algoritmos

Estos algoritmos permiten completar el vector de datos o campos necesarios en los m nuevos registros de transmisiones almacenadas como resultado del proceso ETL, todas las transmisiones nuevas son únicas y deben ser asignadas a un viaje previamente registrado en el sistema. Para lograr integrar el vector final, se requiere realizar una clasificación, a partir del vector de datos necesario para aplicar las reglas de decisión para cada registro de transmisión obtenido en la fase de preparación de los datos con la ayuda del *algoritmo NN*.

- ❖ Algoritmo NN (Distancia Mínima)

El objetivo principal al emplear este algoritmo es encontrar y asignar el geopunto más cercano a cada una de las transmisiones almacenadas por el proceso ETL; este algoritmo, a pesar de no poseer un nivel de complejidad muy alto, exige un conjunto de entrenamiento muy grande para determinar y asignar eficientemente los valores discretos para ciertos campos en base a los atributos del geopunto asignado para cada una de las transmisiones pertenecientes a un determinado desplazamiento vehicular; así mismo reduce la incertidumbre respecto al tipo de espacio o geolocalización instantánea correspondiente a cada transmisión. En la fig. 3, se puede apreciar el pseudocódigo del *algoritmo NN* necesario para obtener los campos requeridos a partir del geopunto más cercano.

- ❖ Algoritmo ID3 (Reglas de decisión)

La finalidad de este algoritmo es: clasificar, predecir o determinar las regiones críticas existentes (incidentes) asociadas al campo “*riesgo*” (+ ó -) que corresponde al nivel de riesgo existente para cada una de las transmisiones almacenadas y vinculadas a un determinado viaje registrado en el sistema propuesto. La “*estrategia de aprendizaje con el algoritmo ID3*”, es respaldada a partir de los ejemplos que se

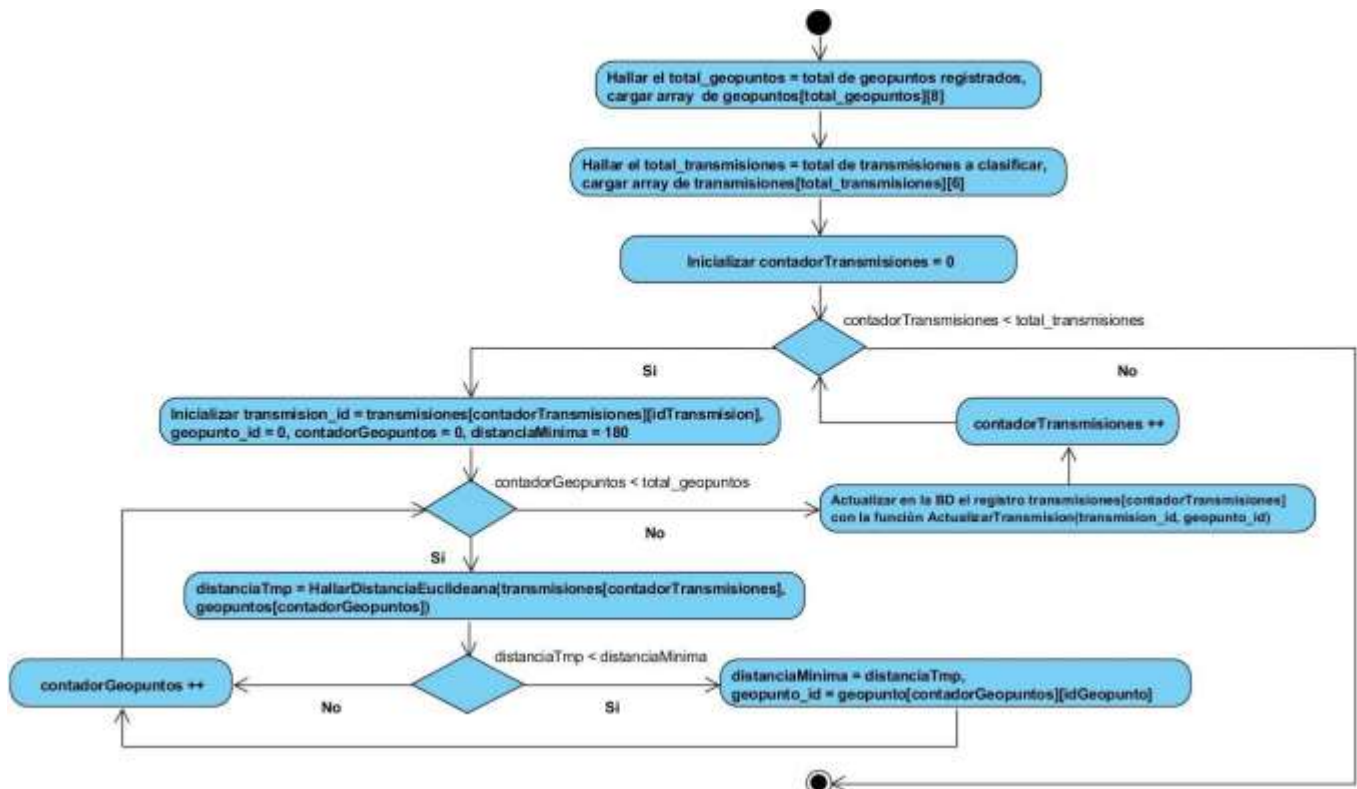


Figura 3. Pseudocódigo del algoritmo NN.

describen mediante un conjunto de atributos, que en un momento dado toman unos valores concretos. En la fig. 4 se muestra un conjunto de atributos y sus valores según los tipos de riesgo para su clasificación como región crítica o región no crítica. Las líneas continuas indican que el valor del atributo se corresponde con la clase nivel de riesgo dada, a diferencia de las discontinuas que expresan no coincidencia.

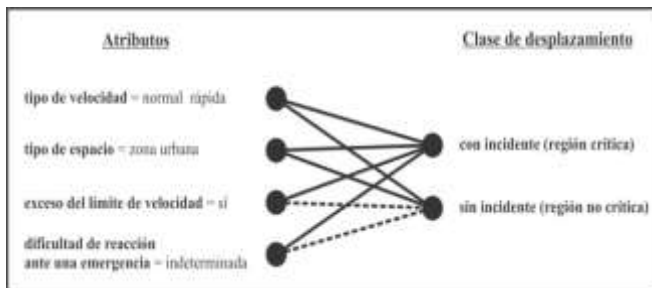


Fig. 4. Atributos y clases.

Del ejemplo de la fig. 4, se puede inferir la siguiente regla:

$$Si (a_1 = val_1) \wedge (a_2 = val_2) \wedge \dots \wedge (a_n = val_n) \text{ entonces } (C_k = val_k)$$

siendo

$$a_1 = \text{tipo de velocidad,}$$

$$a_2 = \text{tipo de espacio,}$$

$$a_3 = \text{exceso de velocidad;}$$

mientras que $c_1 = +$ (región crítica) y $c_2 = -$ (región no crítica).

Para el sistema propuesto, las reglas se generan a partir de los ejemplos o muestras de entrenamiento contenidas en la tabla I.

TABLA I
MUESTRAS DE ENTRENAMIENTO

Muestra	Tipo de velocidad	Tipo de espacio	Exceso del límite de velocidad	Dificultad de reacción ante una emergencia	Clase (región crítica)
1	defensiva	recta	no excede	determinada	NO (sin incidente)
2	defensiva	curva normal	no excede	determinada	NO (sin incidente)
3	defensiva	otro	no excede	determinada	NO (sin incidente)
4	normal rápida	recta	no excede	determinada	NO (sin incidente)
5	normal rápida	curva normal	no excede	determinada	NO (sin incidente)
6	normal rápida	otro	no excede	determinada	NO (sin incidente)
7	normal rápida	recta	si excede	determinada	NO (sin incidente)
8	normal rápida	recta	no excede	indeterminada	NO (sin incidente)
9	normal rápida	curva normal	no excede	indeterminada	NO (sin incidente)
10	normal rápida	otro	no excede	indeterminada	NO (sin incidente)
11	normal rápida	recta	si excede	indeterminada	NO (sin incidente)
12	normal rápida	curva normal	si excede	indeterminada	SI (con incidente)
13	normal rápida	otro	si excede	indeterminada	SI (con incidente)
14	normal	recta	si excede	indeterminada	SI (con incidente)
15	normal	curva normal	si excede	indeterminada	SI (con incidente)
16	normal	otro	si excede	indeterminada	SI (con incidente)

Pajares [14], menciona que para construir el árbol, el orden de los atributos se elige de mayor a menor capacidad discriminante. Una forma de realizar esto es considerar el contenido de información derivado de la teoría de la información. Para ello se definen los siguientes términos:

$$p: \text{porcentaje de ejemplos } + = \frac{|E+|}{|E+| + |E-|}$$

$$n: \text{porcentaje de ejemplos } - = \frac{|E-|}{|E+| + |E-|}$$

siendo $|E+|$ y $|E-|$ respectivamente el número de ejemplos positivos y negativos observado. El contenido de información de un conjunto de datos es:

$$\text{infor}(p, n) = -p \log_2(p) - n \log_2(n)$$

siendo una medida de la entropía (grado de desorden) existente. Este concepto está relacionado con la teoría de información en el sentido de que un evento seguro con $p = 1$ no contiene información (información nula). La información está relacionada con la entropía, a mayor entropía menor información. Por tanto, para construir el árbol hay que elegir los atributos con mayor información (menor entropía), que se mide a través de la función de mérito como sigue:

$$\text{Mérito de un atributo } a_m: \text{mérito}(a_m) = \sum_{i=1}^n r_i \text{infor}(p_i, n_i)$$

siendo:

p_i = % de ejemplos + en la rama i ;

n_i = % de ejemplos - en la rama i ;

$r_i = \frac{p_i + n_i}{N}$ porcentaje de ejemplos en la rama i ,

con N número de ejemplos,

mide la ponderación del grado de desorden de la rama correspondiente con respecto a los ejemplos que contiene. Seguidamente, se construyen las tablas de información tal como se muestra en la tabla 2.

TABLA II
TABLAS DE CONTENIDO DE INFORMACIÓN

Exceso del límite de velocidad	
no excede = 9	si excede = 7
$p_1 = 0/9$ $n_1 = 9/9$	$p_2 = 5/7$ $n_2 = 2/7$

Tipo de velocidad		
defensiva = 3	normal rápida = 10	excesiva = 3
$p_1 = 0/3$ $n_1 = 3/3$	$p_2 = 2/10$ $n_2 = 8/10$	$p_3 = 3/3$ $n_3 = 0/3$

Dificultad de reacción ante una emergencia	
determinada = 7	indeterminada = 9
$p_1 = 0/7$ $n_1 = 7/7$	$p_2 = 5/9$ $n_2 = 4/9$

Tipo de espacio		
recta = 6	curva normal = 5	otro = 5
$p_1 = 1/6$ $n_1 = 5/6$	$p_2 = 2/5$ $n_2 = 3/5$	$p_3 = 2/5$ $n_3 = 3/5$

En consecuencia, el orden de los atributos por su valor informativo es: *exceso de velocidad*, *tipo de velocidad*, *estabilidad* y *finalmente tipo de espacio*.

$\text{mérito}(\text{exceso de velocidad}) = 9/16 \text{ infor}(0/9, 9/9) + 7/16 \text{ infor}(5/7, 2/7)$ $= 0.37761$
$\text{mérito}(\text{tipo de velocidad}) = 3/16 \text{ infor}(0/3, 3/3) + 10/16 \text{ infor}(2/10, 8/10) + 3/16 \text{ infor}(3/3, 0/3)$ $= 0.45121$
$\text{mérito}(\text{dificultad de reacción}) = 7/16 \text{ infor}(0/7, 7/7) + 9/16 \text{ infor}(5/9, 4/9)$ $= 0.55748$
$\text{mérito}(\text{tipo de espacio}) = 6/16 \text{ infor}(1/6, 5/6) + 5/16 \text{ infor}(2/5, 3/5) + 5/16 \text{ infor}(2/5, 3/5)$ $= 0.85060$

Fig. 5. Orden de atributos por mérito.

El árbol de decisión construido en el sistema es como se muestra en la fig. 6, el mismo que representa los datos.

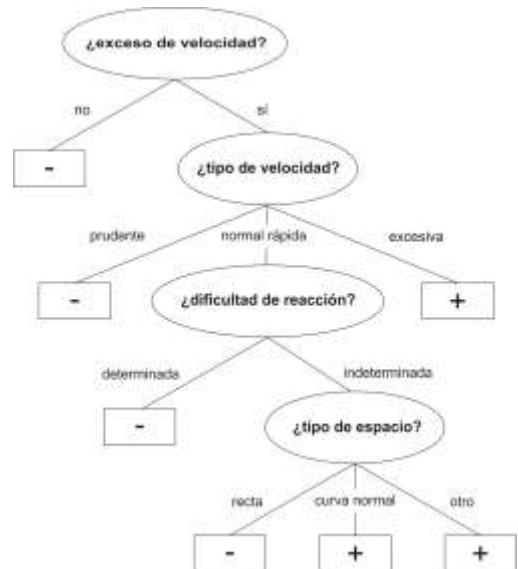


Fig. 6. Árbol de decisión.

El siguiente paso consistió en reflejar este conocimiento, mediante un conjunto de reglas. Básicamente, cada camino desde la raíz hacia una hoja corresponde a una regla, siendo las hojas las representantes de la una clase según corresponda la conjunción de las decisiones asociadas a una rama del árbol. El conjunto de reglas correspondiente al árbol es:

- Si(exceso de velocidad=no) => (clase=-)
- Si(exceso de velocidad=sí) y (tipo de velocidad=prudente) => (clase=-)
- Si(exceso de velocidad=sí) y (tipo de velocidad=excesiva) => (clase=+)
- Si(exceso de velocidad=sí) y (tipo de velocidad=normal rápida) y (dificultad de reacción=sí) => (clase=-)

- Si(exceso de velocidad=sí) y (tipo de velocidad=normal rápida) y (dificultad de reacción=no) y (tipo de espacio=recta) => (clase=-)
- Si(exceso de velocidad=sí) y (tipo de velocidad=normal rápida) y (dificultad de reacción=no) y (tipo de espacio=curva normal) => (clase=+)
- Si(exceso de velocidad=sí) y (tipo de velocidad=normal rápida) y (dificultad de reacción=no) y (tipo de espacio=otro) => (clase=+)
- Análisis, diseño y desarrollo del sistema

Cuando se pretende construir un sistema, es importante considerar una serie de pasos que se deben seguir a fin de obtener un resultado de calidad y que en la Ingeniería de Software se conoce como proceso de desarrollo de software. Para obtener un desarrollo exitoso, resulta necesario definir el marco de trabajo a utilizar, para este caso, se optó por elegir las fases principales del proceso de desarrollo RUP (*Rational Unified Process*). En RUP los cinco flujos de trabajo son: requerimientos, análisis, diseño, implementación y prueba, los cuales tienen lugar sobre las cuatro fases: concepción, elaboración, construcción y transición.

E. Fase 5: Evaluación.

En esta fase se realizó una serie de análisis en base a la validación de los resultados obtenidos a partir de un archivo de prueba; teniendo en cuenta que el nivel de confiabilidad e integridad de los conjuntos de entrenamiento debe ser muy alto, debido a que éstos cumplen un rol muy importante y crítico al momento de realizar la interpretación de los resultados que pueden generar información equivocada para la toma de decisiones. Por tanto, esta fase está sujeta al análisis y reevaluación más detallada del conjunto de entrenamiento o geopuntos, contemplando así la creación, eliminación y/o actualización de un registro en el conjunto de entrenamiento. Esta fase de evaluación, se apoya en la retroalimentación y puede caracterizarse como cíclica, ya que un sistema o aplicación siempre está susceptible a ser mejorado.

F. Fase 6: Despliegue.

En esta fase se procedió a implementar el sistema de monitoreo para la visualización de los datos obtenidos en los diferentes procesos de tratamiento de los datos en el camino a la obtención de información concisa que nos permitirá evidenciar si hubo o no riesgo en el desplazamiento de un vehículo. Así, es posible obtener una mejor interpretación respecto a los reportes clásicos usados hasta ahora en la plataforma de monitoreo principal.

V. RESULTADOS

Para evaluar la funcionalidad del sistema propuesto se mostró el funcionamiento del mismo a un conjunto de expertos y usuarios finales de SUTRAN. Posteriormente se procedió a

aplicar una encuesta, la misma que fue respondida por cada uno de los expertos y usuarios consultados. El 80% de estos expertos y usuarios consideraron que el nuevo flujo propuesto en el sistema para revisar el desplazamiento histórico de un vehículo a nivel de viajes es el adecuado, por cuanto la información desplegada por el sistema propuesto, en especial con la determinación de las regiones críticas en una ruta específica, les permitirá a las empresas de transporte terrestre efectuar una mejor toma de decisiones y acciones a corto, mediano y largo plazo.

En la fig. 7 se muestran los resultados obtenidos con el sistema de monitoreo satelital vigente del viaje realizado por un vehículo de transporte terrestre entre las ciudades de Puno a Arequipa, la distancia recorrida fue de 312 Km. y el número de transmisiones del GPS a la SUTRAN fueron 575. Y la fig. 8 el detalle de los eventos registrados por el sistema vigente.



Fig. 7. Resultados caso de prueba con sistema vigente



Fig. 8. Detalle de la transmisión

Para esta prueba se registraron los eventos de la Tabla 3 en las 575 transmisiones.

Como se puede observar el sistema actual considera que las 575 transmisiones son válidas lo cual no necesariamente es cierto. Del mismo modo, no es capaz de determinar las regiones críticas a lo largo del trayecto de la ruta, por lo que

este sistema ofrece cierto tipo de información que es insuficiente para las necesidades actuales de las empresas.

TABLA III
EVENTOS REGISTRADOS PARA EL CASO DE PRUEBA CON EL SISTEMA VIGENTE

Botón de pánico	Parada no autorizada	Exceso de velocidad	Desvío de ruta
0	15	34	0

Tipo de espacio	Total de transmisiones
Recta	575
Curva Normal	0
Otro (curva peligrosa o zona urbana)	0

En la fig. 9, se muestran los resultados obtenidos con el sistema propuesto al mismo caso de prueba. Se puede observar que el sistema tiene la posibilidad de determinar las regiones críticas (en rojo), además de precisar cuál es el escenario (rectas, curvas normales, curvas peligrosas o zonas urbanas) en el que la región crítica se encuentra. Es decir, con esta información les será mucho más factible la posibilidad de prever acciones a nivel estratégico que conlleve la reducción de costos, un mejor servicio y una mayor rentabilidad. En la fig. 10 se muestra el detalle de los eventos registrados en el sistema propuesto.

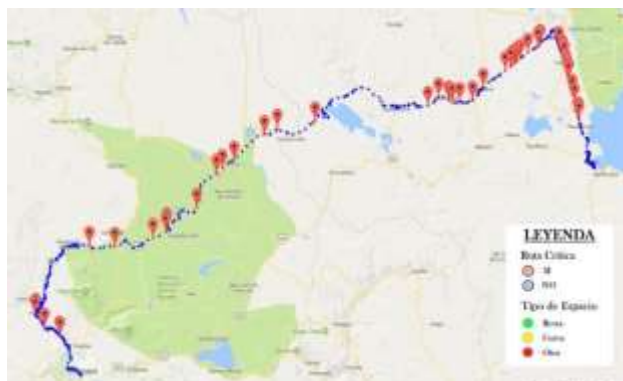


Fig. 9. Resultados del caso de prueba con el sistema propuesto



Fig. 10. Detalle de los eventos registrados en el sistema propuesto

Del mismo modo la información obtenida por el sistema propuesto para el caso de prueba fue la que se muestra en la

tabla 4, en la que se observa que el sistema consideró 561 transmisiones válidas de las 575.

TABLA III
INFORMACIÓN OBTENIDA PARA EL CASO DE PRUEBA CON EL SISTEMA PROPUESTO

Regiones críticas en rectas	Regiones críticas en curvas normales	Regiones críticas en curvas peligrosas o zonas urbanas
33	2	19

Tipo de espacio	Total de transmisiones
Recta	226
Curva Normal	48
Otro (curva peligrosa o zona urbana)	287 (255 Zona Urbana)

VI. CONCLUSIONES

En el presente trabajo se ha descrito el proceso de implementación de una solución tecnológica a un problema real que tienen cientos de empresas de transporte terrestre en el Perú, la misma que consistió en el desarrollo de un sistema de monitoreo satelital para la determinación de regiones críticas en el desplazamiento vehicular a través de una ruta específica haciendo uso de técnicas de aprendizaje supervisado.

Del mismo modo, la efectividad del sistema propuesto se ve reflejada en la obtención del nuevo conocimiento útil, considerando además, que la funcionalidad para personalizar los límites de velocidad permisible para una determinada región o geopunto es una necesidad de los usuarios finales.

El sistema propuesto, puede ser utilizado como apoyo a la toma de decisiones, puesto que genera un valor agregado a los almacenes de datos de los sistemas de monitoreo vehicular satelital utilizados en los terminales terrestres.

REFERENCIAS

- [1] F. García, D. Escobar and L. Vásquez, "Minería de datos para la determinación de variables de tránsito mediante la aplicación de monitoreo satelital", Universidad Nacional de Colombia, Colombia, 2010.
- [2] D. Sarmiento, "Prototipo de sistema de información geográfico en entorno web para monitorear eventos en tiempo real de vehículos con GPS aplicando tecnologías de cache de mapas, AJAX y Base de Datos espaciales", Universidad Católica de Santa María, Arequipa, Perú, 2012.
- [3] M. Basulto, "Descubrimiento de conocimiento sobre accidentes de tránsito en una base de datos concierne a las afectaciones a la infraestructura de las telecomunicaciones en ETECSA", Instituto de Cibernética, Matemática y Física, La Habana, Cuba, 2013.
- [4] E. Huerta, A. Mangiaterra y G. Noguera, "GPS: Posicionamiento satelital", UNR Editora, 2005, pp. 1-14.
- [5] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer y R. Wirth, "CRISP-DM 1.0 Step-by-step data mining", 2000.
- [6] Decreto Supremo N° 017-2009-MTC, "Reglamento Nacional de Administración de Transporte", 2009.
http://transparencia.mtc.gob.pe/idm_docs/normas_legales/1_0_2789.pdf
- [7] Accidentes Declarados en las Unidades de la PNP, "Accidentes de tránsito fatales y no fatales por año, según causa: 2006-2015", 20015.

- http://www.mtc.gob.pe/estadisticas/files/cuadros/Transportes_Carretero_2_5_2.xlsx
- [8] Centro de Gestión y Monitoreo, 2017.
<http://www.sutran.gob.pe/monitoreo-gps/>
- [9] Reporte de Tracking de Unidades, 2015.
<http://www.uvicar.com.pe/servicios>
- [10] Decreto Supremo N° 016-2009-MTC, “Texto Único Ordenado del Reglamento Nacional de Tránsito”, 2009.
http://transparencia.mtc.gob.pe/idm_docs/normas_legales/1_0_1669.pdf
- [11] Informe N° 224-2013/GEL – INDECOPI, “Informe sobre Proyecto de Ley N° 2799/2013-CR”, 2013.
<https://www.indecopi.gob.pe/documents/20182/178372/INF-224-2013-GEL.pdf/14557f10-bb2e-4ed9-9b49-d104a010044b>
- [12] Decreto Supremo N° 009-2015-MTC, “Aprueban modificaciones al Reglamento Nacional de Administración de Transporte, al Texto Único Ordenado del Reglamento Nacional de Tránsito, al Reglamento Nacional de Inspecciones Técnicas Vehiculares, al Reglamento Nacional de Licencias de Conducir Vehículos Automotores y No Motorizados de Transporte Terrestre y al Reglamento de Placa Única Nacional de Rodaje”, 2015.
http://transparencia.mtc.gob.pe/idm_docs/normas_legales/1_0_3611.pdf
- [13] Consejo Nacional de Seguridad Vial, “Plan Nacional de Seguridad Vial 2015-2024”, 2015.
https://www.mtc.gob.pe/cnsv/Proyecto%20del%20Plan%20Nacional%20de%20Seguridad%20Vial%202015_2024.pdf
- [14] G. Pajares y M. Santos, “Inteligencia Artificial e Ingeniería del Conocimiento”, Alfaomega, Mexico, 2006, pp 217-231.
- [15] Ley N° 29380, “Ley de creación de la Superintendencia de Transporte Terrestre de Personas, Carga y Mercancías (SUTRAN)”, 2009.
http://www.sutran.gob.pe/wp-content/uploads/2015/08/Ley_29380.pdf